2024 Edition

Phenotype Phebruary

February 20th, 2024
Community call update
Week 1: Alzheimer's disease (AD)
- AD Literature review complete
- AD definition replicated
- Community call update
- Phenotype Representation analysis for AD
- Cohort Diagnostics PheValuator run
- Atlas demo

Week 2: Non-Small Cell and Small Cell Lung Cancer (NSCLC)
- LC Literature review complete
- Community call update
- NSCLC definition replication
- A discussion on what can be replicated

Week 3: Major Depressive Disorder
- MDD Literature review complete
- Community call update
- MDD definition replication
- PAH data extr. in progress
- Atlas QA

Week 4: Pulmonary Arterial Hypertension
- Community call update
W1: Alzheimer’s disease
Where we are with Alzheimer’s

Reviewed the literature -> Replicated the cohorts -> Characterized the patients -> Estimated performing characteristics of the definition

What else can we learn?
Measurement error impact on background incidence

Inputs
- 13 AD definitions
- 7 databases
- SN, SP, PPV, NPV
  - Joel Swerdel
- Background IR/1000PY
- Errors and IRs age × sex stratified

Impact evaluation
- Correct IR via QBA principles

\[
\text{Outcomes}_{\text{Corrected}} = \frac{\text{Outcomes} - (1 - \text{SP}) \times \text{Persons}_{\text{Al-risk}}}{(\text{SN} - (1 - \text{SP}))}
\]

- Metrics
  - Relative IR
  - Expected absolute measurement error: \( \text{abs}(\log(\text{relative IR})) \)

AD: Alzheimer’s disease, SN: sensitivity, SP: specificity, PPV: positive predictive value, NPV: negative predictive value
IR/1000PY: incidence rate per 1000 person-years, QBA: quantitative bias analysis
### Measurement error impact on background incidence

Harris defn: [2 Dx] OR [2 Rx] OR [1Dx AND 1Rx]; 2\textsuperscript{nd} event [1-365d]

<table>
<thead>
<tr>
<th>Method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sensSpec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Database</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>truven_mdcg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strata</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Female &gt;=85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Persons on 1 Jan 2018-2022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer's disease per Harris JAD 2023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>databaseid</th>
<th>oName</th>
<th>stratum</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>truven_mdcg</td>
<td>Alzheimer's disease per Harris JAD 2023</td>
<td>Female &gt;=85</td>
<td>888</td>
<td>56271</td>
<td>414</td>
<td>368</td>
</tr>
</tbody>
</table>

**sens**: sensitivity, **spec**: specificity, **ppv**: positive predictive value, **npv**: negative predictive value

**IR**: incidence rate/1000 person-years, **cIR**: corrected IR, **IRrel**: relative IR, **IReame**: expected absolute measurement error

<table>
<thead>
<tr>
<th>databaseid</th>
<th>method</th>
<th>tName</th>
<th>oName</th>
<th>stratum</th>
<th>IP</th>
<th>cIP</th>
<th>IPrel</th>
<th>IPeame</th>
</tr>
</thead>
<tbody>
<tr>
<td>truven_mdcg</td>
<td>sensSpec</td>
<td>Persons on 1 Jan 2018-2022</td>
<td>Alzheimer's disease per Harris JAD 2023</td>
<td>Female &gt;=85</td>
<td>53.8824</td>
<td>66.6329</td>
<td>1.237</td>
<td>0.212</td>
</tr>
</tbody>
</table>

**IP**: /1000 P, **IR**: /1000 PYs
Measurement error impact on background incidence

Package update:

- [https://github.com/ohdsi-studies/PhePheb2024](https://github.com/ohdsi-studies/PhePheb2024)
- Development near complete, thanks Thomas Falconer
- 4 data partners signed up to execute ❤️
W3: Major Depressive Disorder
What did we do?

- 497 manuscripts identified for review (2020 - 2024)
- 109 manuscripts filtered based on automated scoring mechanism
  - Prospective studies
  - No full text
  - No detail on the phenotypes
- Data extraction performed on 24 manuscripts

Thank you:
Hayden Spence, Thamir Alshammari, Atif Adam, Jessica Mo, Bill Baumgartner, Buchi Anikpezie, Ruochong Fan, Septi Melisa
What did we do?

497 manuscripts identified for review (2020 - 2024)

109 manuscripts filtered based on automated scoring mechanism*

Data extraction performed on 24 manuscripts

- Prospective studies
- No full text
- No detail on the phenotypes

*Opportunities for automated/systematic lit review to support phenotype development
Do researchers aim at reproducibility or conceptual definitions?

Conceptual definitions: provide some rationale for population chosen and how the criteria selected facilitate capture of such population

Good ex.:
“...To avoid potential bias from other neuropsychiatric conditions, patients were excluded if they had diagnoses for bipolar/manic disorder, mood disorders other than MDD, Alzheimer disease, Parkinson disease, or dementia during the study period...” [aim: cost and utilization]

“...We excluded hospitals wards with fewer than 20 recorded admissions ... this exclusion was made because hospital wards with only sporadic admissions were potentially less likely to report data to the Danish Depression Database because of possible inadequate routines...” [aim: quality of care]

Bad ex.:
Say nothing but exclude codes like F32.5 Major depressive disorder, single episode, in full remission or F32.8 Other depressive episodes
Do researchers aim at reproducibility or conceptual definitions?

Reproducibility

# of papers that have codes: 21/24 (*17 put the codes in the body of manuscript*)

2 papers had codes in supplements, but supplements are not accessible
1 paper does not have codes at all

*The study population consisted of adults with an episode of depression during the study period with no prior antipsychotic use and no prior diagnosis of bipolar disorder or schizophrenia. Episodes were defined by the prescription of antidepressants and presence of depression diagnoses identified with Read codes. Antidepressant prescriptions were grouped into spells of treatment, separated by gaps*

# of papers that explicitly state codes: 5/24 (4 in body and 1 in supplements)*

*Does not imply that definitions are reproducible*
Do researchers aim at reproducibility or conceptual definitions?

Reproducibility: OHDSI studies

1 study provided explicit list of ICD10CM and ICD9CM codes (US data sources only)

1 study provided SNOMED ancestor (US + Korea)
Glance at phenotype definitions

Common patterns in concept sets:
- all F32 (Depressive episode) and F33 (recurrent MDD) and/or corresponding ICD9CM
- F33 only
- F32 and F33 excluding codes that mention remission

Common patterns in phenotype definitions:
- 1/2/3 codes with various restrictions (time window, position, etc.)
- exclusion of differential diagnoses (bipolar, psychosis, dementia, etc.)

More details when we replicate the cohorts!
Next steps

• PAH data extraction in progress (BIG THANK YOU)
• MDD cohort replication (sign up in the sheet)
• PAH cohort replication will follow
• Study package (Jamie in contact with data partners)
• Open call to plan the manuscript